

Modeling the Temporal Dynamics of IT Neurons in Visual Search: A Mechanism for Top-Down Selective Attention

Marius Usher

Carnegie Mellon University

Ernst Niebur

California Institute of Technology

Abstract

■ We propose a neural model for object-oriented attention in which various visual stimuli (shapes, colors, letters, etc.) are represented by competing, mutually inhibitory, cell assemblies. The model's response to a sequence of cue and target stimuli mimics the neural responses in infero temporal (IT) visual cortex of monkeys performing a visual search task: enhanced response during the display of the stimulus, which decays but remains above a spontaneous rate after the cue disappears. When, subsequently, a display consisting of the target and several distractors is presented, the activity of all stimulus-

driven cells is initially enhanced. After a short period of time, however, the activity of the cell assembly representing the cue stimulus is enhanced while the activity of the distractors decays because of mutual competition and a small top-down "expectational" input. The model fits the measured delayed activity in IT-cortex, recently reported by Chelazzi, Miller, Duncan, and Desimone (1993a), and we suggest that such a process, which is largely independent of the number of distractors, may be used by the visual system for selecting an expected target (appearing at an uncertain location) among distractors. ■

INTRODUCTION

Visual selective attention has been the subject of intensive research and debate in both behavioral vision research and in physiology. A first distinction in the analysis of visual information processing is usually made between attentive and preattentive mechanisms (Neisser, 1967; Julesz, 1986; Treisman & Gelade, 1980). While preattentive processes are automatic, effortless, and parallel, and carry out computations involving detection of feature gradients such as texture boundaries or pop-out, attentive processes are thought to involve some "mental effort" and to be required for the identification of objects. The latter are generally correlated with seriality,¹ reflecting a capacity limitation (Bergen & Julesz, 1983; Sagi & Julesz, 1985, 1987; Treisman & Gelade, 1980). In particular, by using a dual report method, Braun and Sagi (1990, 1991) have shown that even simple visual discrimination tasks interfere with each other when they are performed simultaneously during a short time window, thus implying an attentional capacity limitation. Other tasks, such as feature gradient detection, do not engage the attentional system, as they can be performed simultaneously and without interfering with a discrimination task (or with other detection tasks).

Visual attention is further analyzed in terms of two

seemingly opposite concepts. The first is a *focal* serial process, responsible for fast scans of visual displays at rates that do not allow eye movements (30–60 msec/item). This concept of focal attention, going back to Helmholtz (1867), is mostly visualized in terms of a *spotlight* (Treisman, 1982; Crick, 1984) or *zoom lens* (Eriksen & St. James, 1986), where stimuli are processed in higher detail than elsewhere in the visual field (Eriksen & Hoffman, 1973). It has been shown that orienting focal attention leads to an improvement in visual performance (Posner, 1980), and it was proposed that it is required for the conjunction of features (Treisman & Gelade, 1980; Treisman & Schmidt, 1982) or for the scrutiny of targets composed of indistinguishable textures (such as "T"s and "L"s; Bergen & Julesz, 1983).

A second concept of visual attention originates from James (1890), who generalized Helmholtz' attentional theory. James juxtaposed focused attention with its opposite, which he called dispersed attention. He postulated that focused and dispersed attention are extremes in the spectrum of attentional states, and he discussed the conditions under which it is possible to spread one's attention over more than one object at a time. Later, his realization that attention is a capacity-limited process that allocates perceptual resources with variable priorities to different parts of a visual display was analyzed

in a more formal way by Shaw and Shaw (1977) and by Shaw (1978). Accordingly, one can search for a target in parallel over a visual display, without having to perform an explicit serial scan.

The polarity between these two concepts of visual attention generated a very intensive debate. Some visual search experiments indicated that reaction times and error rates in tasks where targets are distinguishable from distractors by unique features (as opposed to conjunction tasks) do not depend on the number of distractors (implying a spread attentional process; Duncan, 1989; Jonides, 1980, 1983; Eriksen & St. James, 1986; Treisman & Gelade, 1980; Treisman & Gormican, 1988). On the other hand, other data showed some dependence on the number of distractors even in the simplest orientation discrimination tasks (for sparse displays; Sagi & Julesz, 1987; Palmer, Ames, & Lindsey, 1993; Palmer 1994). As demonstrated by Palmer, these set-size effects are probably due to attentional effects on the decision process, and not on the perceptual process itself. Moreover, other experiments showed that even tasks that were originally thought of as being prototypical for the seriality paradigm, such as searching for conjunctions of features or discrimination between "T" and "L," have a rather pronounced parallel component² (Krose & Julesz, 1989; Treisman & Sato, 1990; Wolfe, Cave, & Franzel, 1989).

On the basis of visual search experiments using spatial cues, Jonides (1980, 1983) proposed that visual attention can function in two distinct modes, one focal and serial and the other spread over space and parallel. While subjects tend to use the focal mode when spatial information regarding the location of a target is available, for many tasks they can also operate in a parallel mode with resources spread equally over the visual field (Eriksen & Yeh, 1985; Jonides, 1983). From a more general perspective, the two attentional modes can be understood as instantiations of a common principle. In both cases, a selection is made: in the focal mode, the selection operates in the space dimension, but in the spread mode, it is *focused* in a feature dimension (e.g., on "red" or an "A"-shaped stimulus) instead of in space. Such attention on feature dimensions reflects a *top down* expectation; the subject actively searches for a specific feature, just as it is the case for spatially defined focal attention in situations when information about the location of stimuli is available. The view that the fundamental process underlying visual attention reflects a selection (in both attentional modes) has been proposed by Duncan (1980) and Duncan and Humphreys (1989). Moreover, Vecera and Farah (1994) have recently shown that either object-based or location-based attention effects can be obtained, depending on the characteristics of the task. According to Duncan's "late selection" theory, the nature of attentional capacity limitations is a bottleneck in the transfer of visual information to short-term visual memory. In a series of experiments, Duncan (1980) showed

that attentional interference is much more severe for displays with multiple targets than for displays showing a single target and multiple distractors. According to late selection theory, distractors are easily discarded in parallel, but targets have to be processed serially in order to reach short-term visual memory.

The task of fully proving seriality or parallel processing in the search for a single target may be intractable on the basis of behavioral experiments alone. As pointed out by Townsend (1972), capacity-limited parallel processes can produce response times that show a linear increase with the number of elements, as well as a 2:1 ratio between non-target and target slopes (as obtained in some search tasks; Treisman & Gelade, 1980). The two modes can in principle be deconfounded by combining behavioral experiments with physiological recordings. However, as we shall discuss in the next section, multi-electrode recordings during visual search tasks (and without eye movements) seem to be essential for this. No such experiments have yet been performed. Furthermore, it would be desirable that future experiments will be guided by theoretical insight from modeling studies, especially in this field where experiments are costly and difficult. In view of this, we believe that an investigation based on a biologically plausible neural model, which examines schemes and limitations of spatially spread feature-oriented visual attention, is especially needed. In the following section we review the results of two recent experiments in physiological psychophysics, which seem to indicate such a visual attention mechanism. Subsequently, we present a neural model for parallel feature-oriented visual attention, and we discuss it in the light of psychophysical and physiological experiments.

BEHAVIORAL PHYSIOLOGY

Two recent experiments provide evidence for parallel processing across space in tasks where monkeys have to search for a target (Chelazzi et al., 1993a; Motter, 1994a, 1994b). In the experiment of Chelazzi et al., monkeys are first presented with a target that is characterized by its shape. Following a delay period of 1.5 sec during which the screen remains blank, a display showing simultaneously the target shape and a distractor shape are presented (in positive trials; in negative trials, only a distractor appears). The monkey has to initiate an eye movement toward the position of the target shape. The response of neurons from inferotemporal (IT) cortical areas, which are known to be active in such short-term memory tasks (Fuster & Jervey, 1980) is recorded. For each neuron, two shape stimuli are selected: a highly effective stimulus, whose presence enhances that neuron's discharge rate, and a "poor" stimulus, which suppresses that neuron's response (but which presumably enhances the response of other neurons). In some trials the target is the effective stimulus and the distractor is the poor stimulus, while in the other trials the opposite

is the case. In the final display, both shapes (the effective and the poor stimulus) are presented together. In both cases, following the presentation of the two-shape display, a short period occurs during which the neuron's response is enhanced. Following this period, however, its activity level remains high only if the target is the neuron's effective stimulus, and it is suppressed when the target is the poor stimulus. The suppression is initiated about 100 msec before the onset of the saccade (see Fig. 4A).

As it seems, the experiments support an interpretation according to which the response of the IT neurons to the second display, showing both shapes, is structured in two phases. First, there is a parallel phase in which the response of the neuron is enhanced due to the presence of the effective stimulus in the display. During this phase, the response does not depend on whether the target is the effective stimulus for the neuron. After this phase, however, the response shows a clear attentional, expectation-driven modulation; the response is enhanced only if its effective stimulus is the target (the stimulus presented before the delay), and is suppressed otherwise.

If, as the serial search hypothesis assumes, the monkey performs a random serial search (targets and distractors appear at randomly selected locations and no information about the target's location is given), one would expect that in some trials the response latency will be shorter for neurons which prefer the target stimulus, while in other trials (when the search begins with the distractor) the neurons that prefer the distractor respond first. Since only the activity of one neuron is recorded at any time, serial search cannot be excluded with absolute certainty. However, when measuring from a single neuron and changing the target, one may still expect a bimodal distribution of latencies (corresponding to trials where the search begins with the target or the distractor). Existing data do not confirm the existence of a bimodal latency distribution (R. Desimone, personal communication). An unambiguous confirmation would require multielectrode recordings; the finding of trial-by-trial anticorrelations in the activities or response latencies between neurons that code for target and distractors, respectively, would suggest seriality even in the absence of bimodal latency distributions. Nevertheless, this experimental result seems to support a "late selection" attentional model (Duncan, 1980) in which after a first parallel search a competition is initiated, and only one object (the target) is accepted to visual short-term memory.

The mechanism of global feature-oriented attention is also supported by experiments performed by Motter (1994a, 1994b). In these experiments, Motter recorded neural discharges in cortical area V4 when monkeys performed a delayed match-to-sample task. Monkeys are first shown a cue of a specific color at the fixation point, indicating the color of the target. Then six shapes appear

on the screen on a ring around the fixation point, only one of them being in the receptive field of the cell whose activity is recorded. The cue is *not* in the receptive field of the recorded cell. Three out of the six shapes have the color of the cue (the target) and the other three have a different color (all the same). The six shapes remain on the screen for a variable period of time (1500–2700 msec). Following this period, two of the three target-colored shapes disappear while the third one remains on the screen for an additional 200 msec before the screen is blanked out entirely, and the monkey has to initiate a saccade toward the location of the only remaining target. During the delay period, when the three targets and the three distractors are on the screen, the monkey cannot respond since the information is still ambiguous (knowing the target's color but not yet its location). Motter finds that the activity during this period is enhanced *for all three* stimuli with the target color and that it is suppressed for all three distractors. Again, it seems that the monkey is using the delay period to eliminate all distractors in parallel. Such a strategy is efficient for the task at hand. Psychophysical evidence from a conjunction search experiment also shows that humans are able to discard all irrelevant color stimuli in parallel (Egeth, Virzi, & Garbart, 1984).

As in the Chelazzi et al. (1993a) experiment, only one neuron is recorded at a given time, and thus the possibility of a serial attentional search cannot be fully discarded. However, the body of psychophysical experiments on monkeys and humans makes the existence of a global attentional mechanism, which depends on top-down expectations for object or features highly probable. In the next section we present a neural model for spread attentional selection. The model was mainly devised in order to account for the results of the experiment in IT cortex by Chelazzi et al. (1993a). Subsequently, the implications of the model and its relationship with other psychophysical experiments are discussed. Experimental aspects not studied by Chelazzi et al. but essential in late selection theory are also addressed.

THE MODEL

Due to their large receptive fields, cells in IT cortex respond to specific shapes, basically irrespective of their location in the visual field. In our model, cells that are sensitive for the same feature are connected to each other by excitatory connections, which could have been formed in a previous learning phase by Hebbian interactions among cells responding to a common visual shape (this learning phase is not modeled). Thus, objects (shapes, colors, etc.) are represented by sparsely connected cell assemblies with excitatory connections (in our simulations, each cell is randomly connected to one-fourth of the cells in the same assembly), and we assume furthermore that different cell assemblies are

mutually inhibitory.³ This connectivity is consistent with the results of the IT recordings, in which the response of neurons coding for one shape is enhanced, and the response of neurons that code for a different shape is suppressed. Inhibition is modeled, according to the constraint imposed by Dale's principle, by a different set of cells. Moreover, in contrast to the excitatory connectivity, inhibition is diffuse and nonspecific, accounting for the physiological indication that only excitatory cells are involved in synaptic learning. In our model, the inhibitory neurons are grouped in a common pool of cells, which mediate the lateral competition. A memory model based on this architecture has been analyzed extensively by Horn and Usher (1990) and Hendin, Horn, and Usher (1991).

An important issue in the neural computation and connectionist literature is the nature of the representations underlying information processing (Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986). While a localized scheme where each unit stands for a small neural population could be a reasonable first approximation, a distributed sparse representation, in which some cells participate in the representation of several objects, allows a much higher storage capacity and is better supported by neurophysiological evidence (Georgopoulos, Kettner, & Schwartz, 1988; Heit, Smith, & Halgren, 1988; Miyashita & Chang, 1988; Tanaka, 1991). Moreover, the use of distributed representations is important in particular when the represented objects satisfy a specific similarity structure; it is then expected that the representations of objects with a larger degree of similarity have larger overlap. In this work, we consider the two simplest situations:

- No similarity structure (i.e., all objects are equally similar). In this case we will use nonoverlapping representations for the various objects.
- A one-dimensional similarity continuum with cyclic boundary conditions, as is the case for colors or line orientations. In this case we will assume that the overlap between two objects that are close on the continuum is larger than that between two objects with a greater distance on this continuum (i.e., dissimilar objects).

The model has been implemented at two levels. At a first *micro* level we simulated individual cells, according to an integrate-and-fire scheme. Although this is clearly a simplified model of spiking cells, previous research has shown that it can mimic the behavior of morphologically detailed compartmental models of pyramidal cells with several voltage-dependent currents at the cell body and a passive dendritic tree (Softky & Koch, 1993; Bernander, Koch, & Usher, 1994). At a second *macro* level we modeled mean-field population dynamics (Wilson & Cowan, 1972; Amit & Tsodyks, 1991; Abbott, 1992).

The case of stimuli whose similarity structure spans a

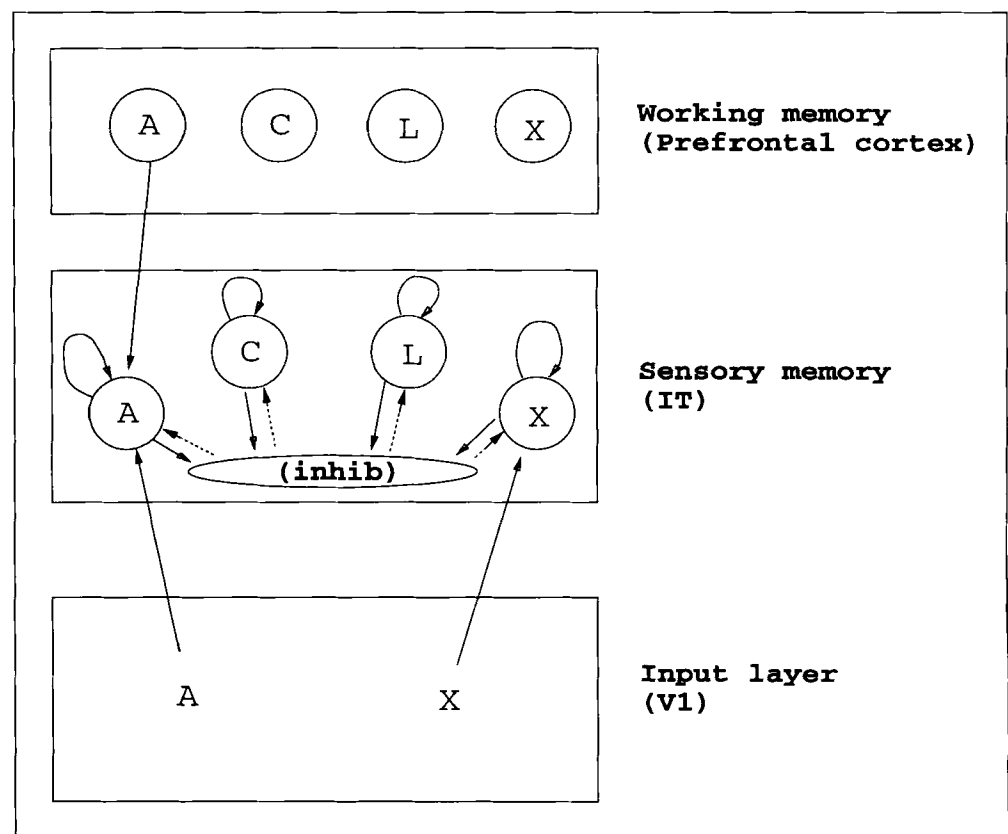
one-dimensional continuum was modeled at the micro-level by a cyclic set of cell assemblies, such that the two cell assemblies representing similar objects, e.g., "yellow" and "orange" color spots share a subset of neurons. At the macro (population) level, where each cell assembly is represented by one variable (the population activation), the overlap of the representation of similar objects is mapped into an excitatory connection that connects "neighboring" assemblies. The strength of this connection is small relative to the strength of self-excitation, which accounts for recurrent connections inside one cell assembly. The exact specification of the dynamics and parameters of both modeling schemes is presented in Appendices A and B.

Each cell assembly receives input from an input layer (e.g., primary visual cortex) where a topographic representation of the visual field may be supposed, but the model does not rely on this, being completely independent of topographic relations. For each object (shape) in the input layer, activation is transmitted to the corresponding cell assemblies. The strength of the input is proportional to the similarity between the "learned" object and the shape present in the display. We assume, however, that it is invariant with respect to transformations such as scale, contrast, and translation. While we do not address here the part of the preprocessing responsible for the scale translation and contrast invariance, we do include an intermediate preprocessing stage that makes the input to each cell assembly in the IT module dependent in a nonmonotonic way on the number of stimuli of that type. The description of this preprocessing stage is given in Appendix C.

In accordance with the results of the IT recordings, the weights of the excitatory connections between cells in each cell assembly are chosen such that they are strong enough to generate (via the inhibitory pool) strong competition between the objects but not strong enough to make the cells' activity remain independent of their input.⁴ Under these conditions, the system is input-sensitive, as found in the Chelazzi et al. data, where the response during the delay is much lower than during the stimulation (7–9 vs. 30–40 Hz) but slightly higher than the spontaneous rate (about 5 Hz; see Fig. 4A). Functionally, we consider this module as a visual sensory memory system.

We assume the existence of an associated working memory module, located presumably in frontal cortex (we do not model the internal dynamics of this module explicitly). This second module has the same architecture as the sensory memory module, with a cell assembly corresponding to each cell assembly in sensory memory. However, the excitation between cells of the same assembly is assumed to be stronger in working memory than in sensory memory.⁵ Consequently, the response of an activated assembly persists in this second module during the delay even in the absence of sensory stimuli (although at a lower level than during sensory stimula-

Figure 1. Architecture of the model. Solid arrows: excitatory connections, dotted arrows: inhibitory connections. Objects "A" and "X" are in the visual field and object "A" is, in addition, stored in the working memory module. See text for a more detailed description.



tion), and is only extinguished at the response onset by a cognitive mechanism that is not part of the present model. This differentiation between the memory characteristics of cells in IT and prefrontal cortex is motivated by the recent finding that, unlike in IT, the delayed activity in prefrontal cortex conveys information about targets even across intervening stimuli (Chelazzi, Miller, Lueschow, & Desimone, 1993b). We further assume that each working-memory cell assembly sends a weak excitatory "top-down" projection to its associate assembly in the sensory memory area. During a search task, when a subject is searching for an expected target, the top-down feedback projection will add a weak additional input to the target assembly in the sensory memory.⁶ This same principle, according to which attention is modeled as an additional bias originating from a working memory module in prefrontal cortex, has been used recently in the context of the Stroop task by Cohen, Dunbar, and McClelland (1990) and by Cohen, Servan-Schreiber, and McClelland (1992). The architecture of our model is displayed in Figure 1.

MODEL BEHAVIOR

Dissimilar Stimuli

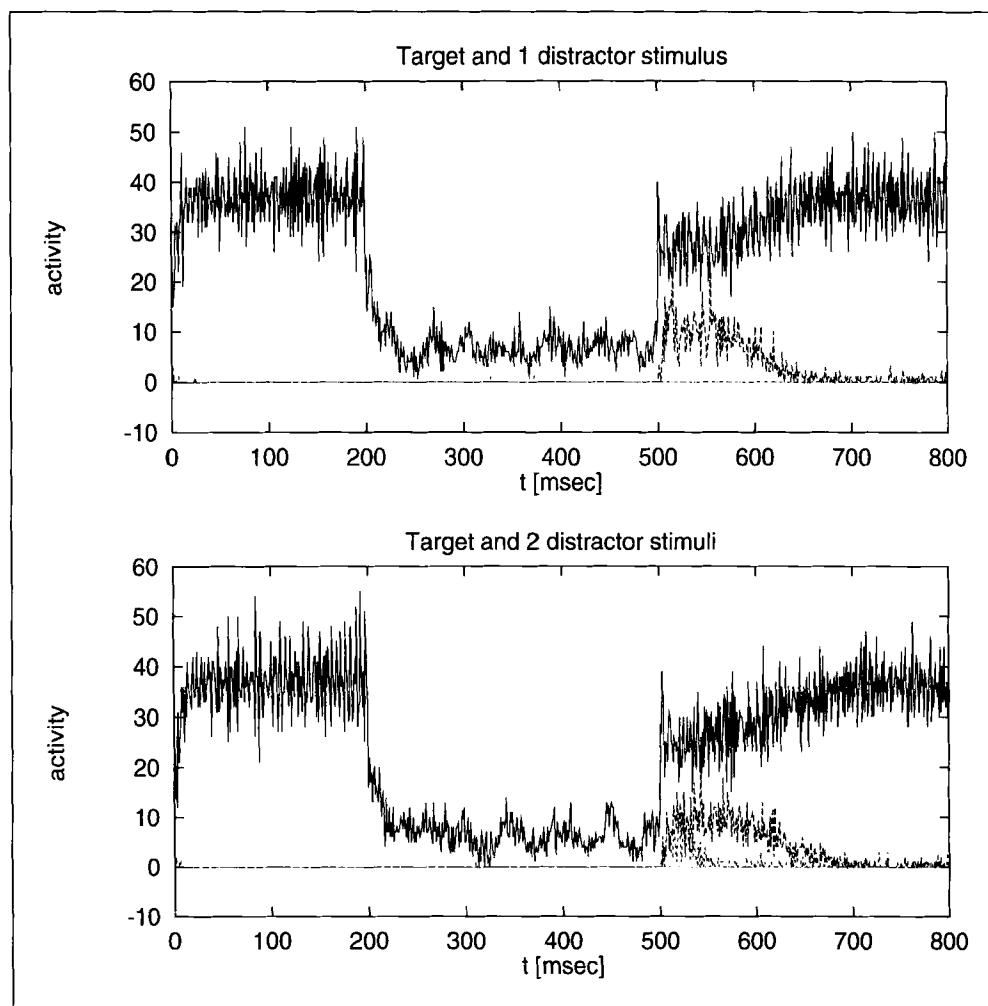
The performance of the model during the delayed match-to-sample task with dissimilar stimuli is described in the following, first for positive trials (target included in the probe display), and then for negative trials.

1. *Positive trials* (the target appears in the final display):

During the presentation of the target (first 200 msec), only one assembly receives input from the input layer. Neurons belonging to this assembly (the "target assembly") are activated at a rate of about 30–40 Hz while the activity of neurons in other cell assemblies is suppressed due to the competitive dynamics (Fig. 2A). The target assembly provides input to the corresponding assembly in the working memory modules. As was mentioned previously, we do not explicitly model the neurons in this module, but rather assume that they remain active during the rest of the task due to stronger connections within each cell assembly, and send a weak top-down projection to the target assembly in the sensory memory.

During the delay, the only specific external input received by the sensory module is this "expectation feedback." We assume that this attentional input is much weaker than the sensory input (a ratio smaller than 0.1 has been used, see Appendices A and B). However, as mentioned above, the attentional input shifts the response curve close to a tangency relationship to the identity (Fig. 8), guaranteeing a very slow decay of the activity during the delay period. During the presentation of the two-shape stimulus, strong input is received in the two corresponding cell assemblies of the sensory memory module. Consequently, the activity of both cell assemblies is initially high [corresponding to the first part of the response in the Chelazzi et al. (1993a) experiment].

Figure 2. Activity of cell assemblies in sensory memory (positive trial). A target is presented during the first 200 msec, and after a delay of 300 msec, a stimulus containing the target and 1 (top) or 2 (bottom) distractors is presented. After a short transient, the target wins the competition and the distractors are suppressed. The simulation shows the activation of cell assemblies corresponding to the target (solid line) and the distractors (dashed and dotted lines) in the integrate-and-fire scheme. The mean-field population dynamics gives similar results (not shown).



This activates the inhibitory assembly, creating strong competition between the two stimulated assemblies. Due to the “top-down” additional input, the target cell assembly wins the competition, and the distractor’s activity is suppressed (Fig. 2).

2. Negative trials (the target does not appear in the final display):

The target and delay period are identical to the positive trial scenario. During the two-shape display, however, the two cell assemblies that receive sensory input are different from the one that receives top-down attentional input. Consequently the target is suppressed and the distractors compete for activation, but without the support of the attentional modulation neither of them completely dominates the module (Fig. 3).

Influence of Target-Distractor Similarity

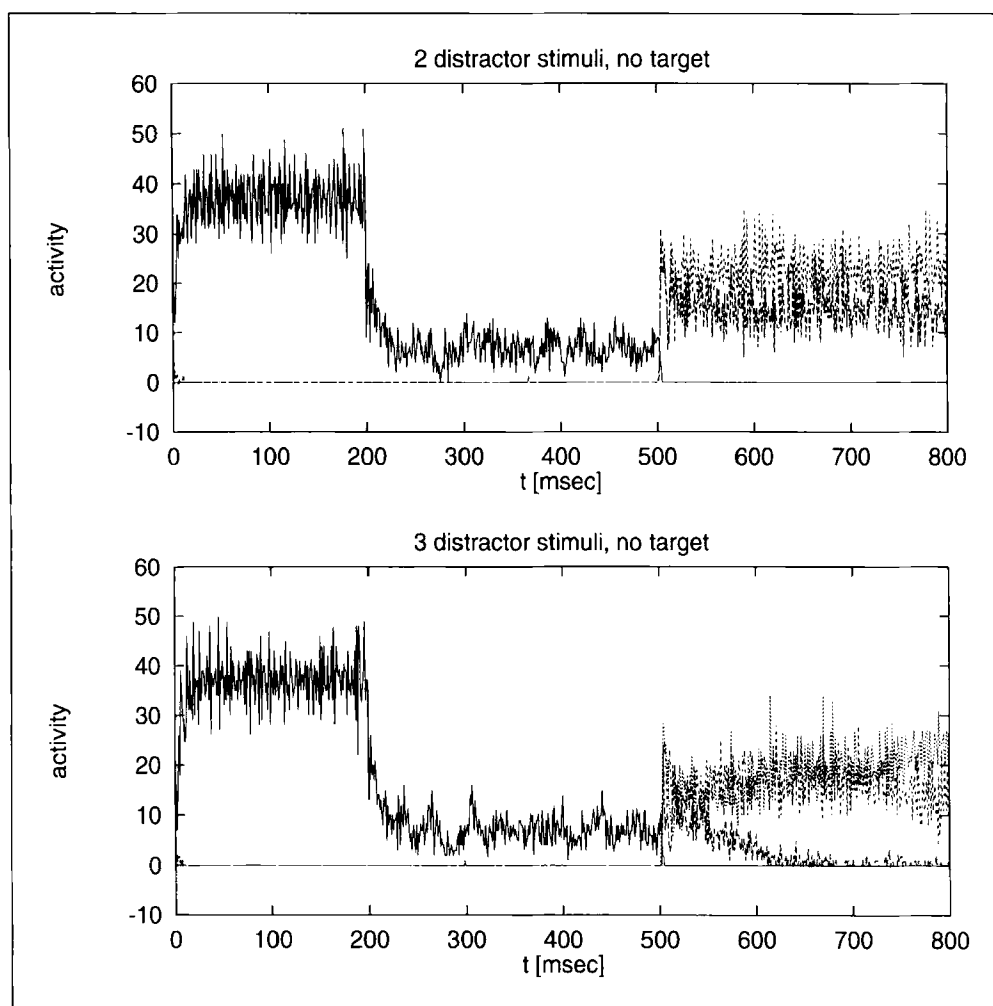
We describe here the network behavior in the model that allows for stimuli that are ordered according to a cyclic similarity structure (as described in the previous section).

Figure 4B and C display the activity in the population

activation scheme for a positive and negative trial, respectively. In this case, the target and the distractors were not similar (they were more than two neighbors away in the ring structure). This case replicates the qualitative features displayed in the situation with dissimilar stimuli (previous section; Fig. 4b and c shows essentially the same phenomenon as Fig. 3 but in the population model). For comparison, we show in Figure 4a the electrophysiological response in the monkey experiment (Chelazzi et al., 1993a).

To see how the behavior is affected by the target-distractor similarity, we simulate the same behavior for a given target, with distractors at different distances from the target (2 or 1 neighboring assemblies away on the ring structure). To show more clearly the similarity effect, we display in Figure 5 the *difference* in activation between the target and each of the distractors (which presumably determines the response) for the three conditions (target-distractor distance on the ring 1, 2, or 3 units). We observe that once the stimulus-distractor display is presented, the dynamic course of the activation is highly sensitive to the stimulus-target similarity. When the stimulus is very different from the distractor (top

Figure 3. Activity of cell assemblies in sensory memory (negative trial). A target is presented during the first 200 msec. After a delay of 300 msec, a stimulus containing only 2 (top) or 3 (bottom) distractors (but no target) is presented. The activity of the cell assemblies corresponding to the target is shown by a solid curve, the activity corresponding to distractors by broken curves. While the activity of the target decays, the distractors continue to compete for activation. Mean-field population dynamics give similar results (not shown).



curve; dotted), the difference between the activation of the target and of the distractor increases fast towards a relatively high level. When the target and distractor are more similar (middle curve; dashed), the difference increases somewhat more slowly and saturates at a slightly smaller asymptotic value. When target and distractor are very similar (bottom curve; solid line) the difference in activation remains low. The reason is that neighboring assemblies send some degree of activation to each other and therefore the competition mediated via the common inhibitory pool is less effective. Consequently, when simultaneously activated by the input, these cell assemblies respond at similar activation levels.

DISCUSSION

We have presented a simple neural model that displays a behavior characteristic of IT cells in delayed match-to-sample tasks. We propose that such a model may shine light on the neural mechanism underlying expectation-driven selective attention for features. The function of the described mechanism is to “look for” an expected stimulus among distractors in the display (the expected

stimulus being the one that had been shown previously, i.e., the “cued” one). When such a target stimulus is found, the corresponding assembly is *selected* for activation, while otherwise (in the absence of the target) no assembly achieves full domination of the system. To further validate the target’s presence, the module may be supplemented by a comparison mechanism. This will elicit a positive response when the most active assembly in the sensory module is the one that corresponds to the assembly in the working memory module that holds the memory of the cue. We note, however, that such an additional mechanism may not be necessary, since a positive response could be initiated on the basis of the maximal activation in the sensory module alone. In this way, instead of comparing the target with the distractors one by one through a serial search, a parallel expectation-driven process selects the most likely stimuli and only one (or no) comparison is required. Although the attentional mechanism proposed is very simple from a functional point of view (no *explicit* matching procedure of bottom-up and top-down information, as in other models), this approach is in the spirit of the connectionist framework that is adopted in approach of our model.

Figure 4. (a) Cell response in monkey IT cortex, as reported by Chelazzi et al. (1993a). Shown is the average firing rate of one cell on trials with either a good (solid line) or a poor (dotted line) stimulus presented as cue. Black horizontal bars on the abscissa indicate presentation of cue and choice stimuli. Average time of saccade onset (297 msec) is indicated by an asterisk. See Chelazzi et al. (1993a) for details. (b) Positive probe: one target and two distractors are presented. The target wins the competition. (c) Negative probe: three distractors are presented. The target is suppressed and the distractors are competing for activation. Simulated using population dynamics, the more detailed integrate-and-fire dynamics gave similar results (not shown). Parts (b) and (c) are shown at a higher temporal resolution than part (a).

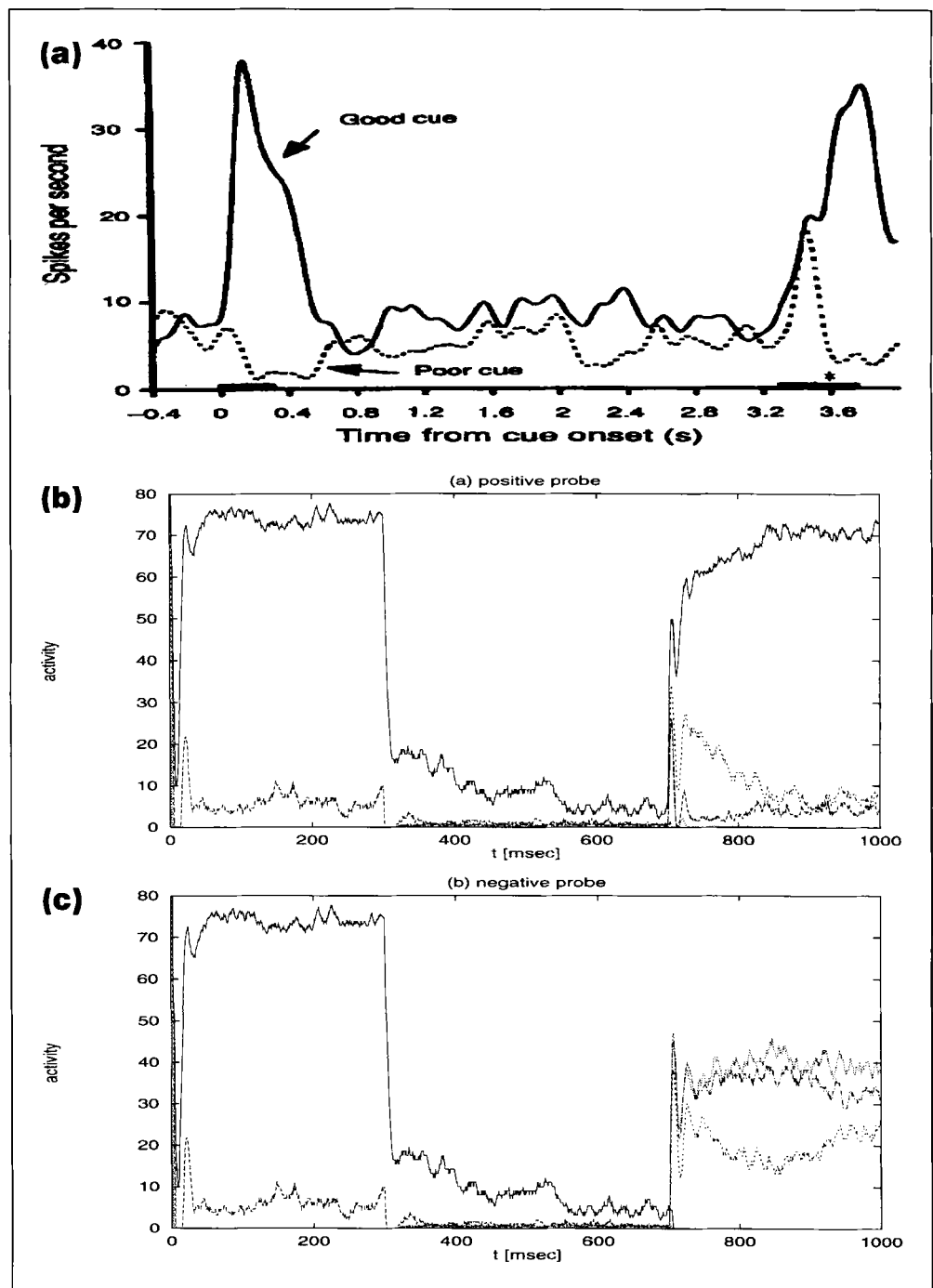
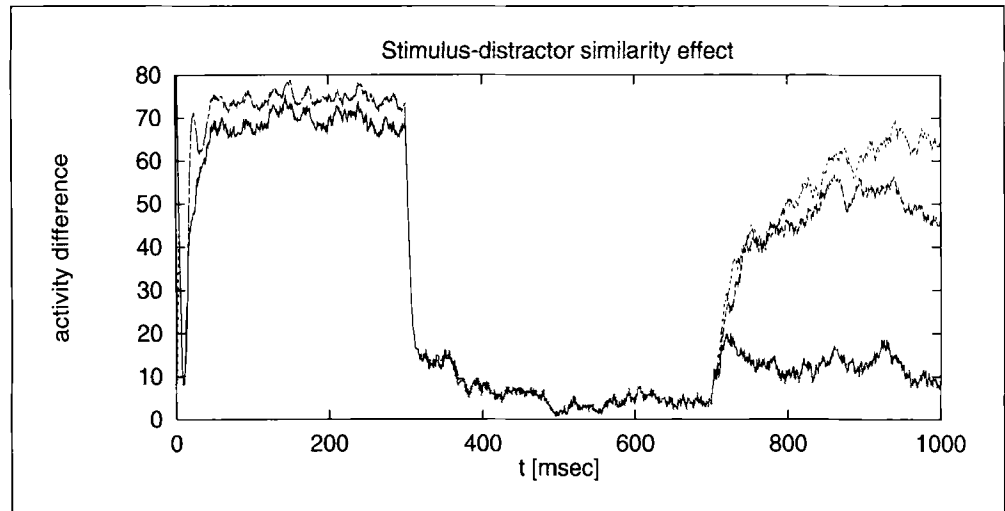


Figure 5. Stimulus similarity effects (population dynamics model). The activation of the target assembly minus the activity of each of the three distractor assemblies is plotted as function of time. The target stimulus is extinguished at $t = 300$ msec and comes back, now together with the distractor stimuli, at $t = 700$ msec. The difference in activity between the target and the distractor increases with their distance from each other. The distance is 1 (solid line, bottom), 2 (dashed line, middle), and 3 (dotted line, upper), respectively.



Once the nature of distributed representations is taken into account, “matching” becomes an *implicit* automatic process.

It may be instructive to compare this mechanism with the assumptions underlying signal detection theory in visual search (e.g., Green & Swets, 1966; Palmer, Ames, & Lindsey, 1993): following presentation of a stimulus, the value of the maximally activated detector (reflecting the similarity of its stimulus to the target) is selected and compared with a threshold, to generate a response. In fact, the mechanism that we propose here can be thought of as neural implementation of a selection mechanism, whose information processing limitations will be discussed below.

We should mention that we do not view this model as revealing the one and only mechanism underlying visual selective attention, but rather consider it as one of several mechanisms, such as focal attention, corresponding to the multitude of physiological structures in visual and associative cortex. Consequently, we expect that most visual tasks will recruit the resources of one or several of these mechanisms, depending on the requirements of the task. Therefore we have not tried to reproduce here all effects found in visual search. Instead, we focused on the delayed match-to-sample paradigm that is accessible to physiological verification (Chelazzi et al., 1993a; Motter, 1994a; Miller, Li, & Desimone, 1993b).

The first and most important point to notice is that the model is not limited to serial processing, which is often assumed to be the case for attentional systems. As observed from Figure 2, the time course of the response to the target stimulus is independent of the number of distractors, as long as the stimuli are well discriminable. In this case, it is as easy for the target to suppress one, two, or more distractors. The reason for this is that all cell assemblies compete for activation; due to the competition between distractors, their presence in a larger number is not detrimental for target activation. This is

very different from searching based on focal attentional mechanisms. While focal attention can search for one spatial position at a time and thus is serial in the number of distractors presented at different locations, the mechanism proposed here searches for one *object* at a time. Consequently it is serial in the number of objects that constitute the response set (targets), but not in the number of distractors. While the mechanism is not serial in the number of distractors, it is nevertheless attentional, insofar as it involves a selection that depends on a top-down expectational input, and insofar as it probably requires an active and voluntary mental effort. Moreover, this mechanism does not contradict the “location dominance” effect found in the experiments of Tsal and Lavie (1988, 1993) since once such a feature selection takes place, it can feed back onto a topographic representation, guiding the attentional spotlight (Cave & Wolfe, 1990), and thus improving visual discrimination. Indeed, this mechanism shows how a target that is not spatially defined can direct the attentional spotlight without having to perform an explicit serial search.

Late Selection and Similarity Effects

The described attentional mechanism may correspond to the late selection theory of Duncan (1980). According to this theory, all distractors are processed preattentively in parallel, and only the selection of the target is serial. This is indeed the case if one considers the first parallel stage of activation of the IT neurons in the Motter and Chelazzi et al. experiments as preattentive, and the late stage when the target comes to dominate the distractors as underlying selective attention. However, according to this assumption, the preattentive and the attentive systems are not two different structures, but rather two dynamic stages of the same module. The way that selection operates is through the interaction between the

bottom-up sensory input supplemented by lateral competition with the top-down attentional bias.

In accordance with Duncan's theory, we find in our model a target-distractor similarity effect; the closer the similarity of target and distractors, the slower and less effective the module becomes in selecting the target. At some point, if the difference between target and distractors becomes too small, a different strategy, such as a focalized serial scan, is required. Alternatively, set size effects are expected due to the fact that noise scales up with the number of distractors. Another similarity effect required by Duncan's late selection theory is the distractor-distractor effect: selection is easier when distractors are similar to each other, which is viewed in his theory as a grouping effect [see also Humphreys & Müller (1993) for a connectionist model of selective attention, which explicitly implements grouping]. While the model presented thus far does not account for this effect by itself, it is highly probable that grouping is part of the preprocessing, which is essential for the functioning of the module. This preprocessing has to account for the fact that the input received by the object assemblies is standardized, so that it does not depend on the size, contrast, brightness, etc., of the presented stimuli. It is also important that a distractor assembly does not receive more input than a target assembly, even though there may be more distractors than targets. While this requirement could seem at first counterintuitive, it is supported by the following physiological and psychophysical evidence.

It has been shown recently that cells in IT cortex respond more strongly to stimuli showing one object than to stimuli showing two objects of the same kind (Miller, Gochin, & Gross, 1993a). While it was not possible to conclude from the experimental evidence whether the effect is global (same response reduction independent of the type of the stimuli) or specific (more reduction for identical or similar stimuli), we should note that a specific inhibitory effect of this kind is exactly what is required for the implementation of a pop-out mechanism: when one target is embedded among many distractors, the distractors will be suppressed more than the target. Similar suppression of stimulus responses by the simultaneous presentation of like stimuli has been demonstrated previously in earlier cortical areas of cat (Blakemore & Tobin, 1972) and monkey (Allman, Miezin, & McGuinness, 1985). Psychophysical evidence for such a suppression is shown by Nakayama and Mackeben (1989) who called it a "similarity vetoing mechanism."⁷ In the context of our model, such preprocessing leads to distractor-distractor similarity effects: several identical (or similar) distractors presented simultaneously in the display provide less input to their associated cell assemblies than the unique target. We implemented an inhibitory preprocessing mechanism with this property based on a scheme for shunting nonlinear inhibition that was proposed by (Abbott, 1991, see Appendix C for details).

Accordingly, similar distractors get less input than dissimilar ones. Indeed, the simulation results (Fig. 6) indicate that the activation difference between the target and the distractors increases faster for similar distractors than for dissimilar ones. Thus the IT model proposed, when supplemented by specific inhibitory preprocessing, satisfies the requirements of late selection theory.

Focal Attention and Conjunction Tasks

If a parallel mechanism for selective attention does, indeed, exist, one may ask whether a focal serial attentional mechanism is still required in search tasks. We suggest that focal serial attention is required in at least two search situations: when spatial information (such as a locational cue) is available, and when the parallel selection mechanism reaches its limitations. One example for such limitations has already been presented, the case of very difficult discrimination between target and distractor stimuli. In such a situation, a different serial mechanism is required, as suggested by the high search slopes obtained in these tasks (Treisman & Gelade, 1980).

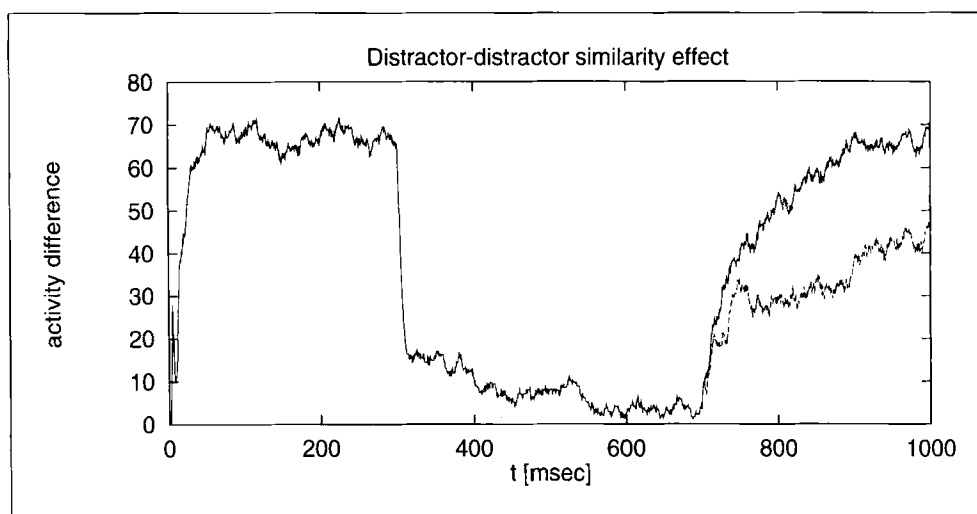
A different limiting situation for the parallel selection mechanism is given by search tasks consisting of conjunctive stimuli (e.g., a target defined as a red vertical line, while distractors are red horizontal and blue vertical lines). In this case the limitation may be due to the nature of the representation. If conjunctive stimuli (e.g., "red-vertical") were represented by Hebbian competing cell assemblies in the memory system, the selection of conjunctive targets would have been as easily accomplished in the model. However, the existence of such conjunctive Hebbian cell assemblies is improbable since the cooccurrence of the two properties (color and orientation in this case) is accidental. At least in the absence of specific training with a set of such stimuli, it is thus more probable that various features (colors, orientations, etc.) are stored in different and independent modules, as postulated by feature integration theory (Treisman & Gelade, 1980).

Assuming that each feature dimension (colors, shapes, etc.) is represented by a different memory module, it is easy to see why the selection mechanism reaches its limitations. In the example presented above (red vertical target among red horizontal and blue vertical distractors), the top-down attentional input is provided to both the red and the vertical assemblies. Therefore, both assemblies will win the interaction whether there is or there is not a conjunctive target in the display. The parallel stage is useless in this case and the decision has to be based on a serial scan of all objects.

Relation to Other Models

Four models accounting for attentive visual processing have been proposed recently (Cave & Wolfe, 1990; Humphreys & Müller, 1993; Cohen et al., 1990, 1992; Gross-

Figure 6. Distractor-distractor similarity effect (population dynamics model). The activation of the target minus the activity of the distractors is plotted as a function of time. The stimulus is applied at $t = 700$ msec. The top curve (solid line) shows that the difference between target activity and distractor activity when the distractors are all identical reaches a higher value. In contrast, when distractors vary from each other, the difference between target activity and distractor activity has a slower time course and reaches a smaller value (lower curve; dashed).



berg, Mingolla, & Ross, 1994). The Humphrey and Müller (1993) model is a bottom-up connectionist implementation of late selection theory that focuses on the role of grouping in the selection process in visual search. A more abstract model that implements a parallel process "guiding" a focal serial search has been proposed by Cave and Wolfe (1990). A functionally very similar model was suggested by Treisman and Sato (1990). In their model, an activation map that combines bottom-up saliency information with top-down expectational input guides a serial search. While their model explains many of the effects found in the visual search literature, they conclude that it is very difficult to distinguish its predictions from Duncan and Humphreys' late selection model, and consequently from the predictions of the model described here. Just as "Guided Search," the model of Grossberg et al. (1994) is capable of reproducing quantitatively a range of experimental results. However, this model is functionally defined and no neural implementation has been developed. In the model by Cohen and co-workers, top-down input from prefrontal cortex is used in the context of the Stroop task. We will look in some more detail at similarities and differences between these four models and our model in the following.

The models by Cave and Wolfe and by Duncan and Müller are different from the model presented here insofar as they use spatially topographic representations. We, on the other hand, have focused on the processing of visual information, which occurs in a less topographically ordered cortical area, IT visual cortex (cells in IT have very large receptive fields and their response is to a large extent translation-invariant). Since earlier cortical areas are more topographically organized, it is possible that a grouping process, such as described in Humphreys and Müller (1993), could be part of the preprocessing of visual input for IT. We have not attempted here to address grouping in detail. The inhibitory mechanism presented above, which can be seen as a simplification of

realistic preprocessing, is, however, sufficient to enable the IT module to display effects observed in the physiological and psychophysical studies.

The Cave and Wolfe (1990) model may correspond to a different mechanism that subserves focal attention. Such a mechanism may be required for tasks in which global computations are not sufficient for reaching a decision. Indeed, the gradual reduction in search slopes for conjunction searches with stimulus saliency and conjunction dimensionality indicates that such a guiding mechanism may indeed be in place in conjunction tasks. The interaction between the different attentional mechanisms, and its role in the choice of a search strategy, is a topic of great significance that should be the subject of further research.

Grossberg et al. (1994) developed a model of attentional search that quantitatively reproduces reaction times for several different visual research experiments. The method is essentially a grouping algorithm, in which all items of the target color are connected by straight lines. Those lines that come too close to items of nontarget colors are eliminated. All connected items are then grouped together, segmenting the visual scene in a number of candidate structures. In the next step, one of these structures is selected randomly and the items in this group are recursively compared with the target. If a match is found, the search task is finished, if not, the next group is tried, until either a match has been found or all possibilities have been exhausted and the target is declared missing. Different from our work, however, their model is defined functionally, in terms of processes that are not immediately realizable in neural hardware. For instance, the first steps in their algorithm are the explicit computation of the equations for the lines connecting items of the target color, and the computation of the distances between their centers. Other steps contain operations like deletion of lines, maintenance and recursive processing of symbolic lists, etc. Grossberg et al. (1994) explicitly state that they specified their work as

an algorithmic model rather than as a neural model (p. 475 in their article).

The fourth model for selective attention, which uses a similar approach to the one presented here, has been proposed by Cohen et al. (1990, 1992) in the context of the Stroop task. As in our model, Cohen et al. assume that attentional modulation occurs through additional input originating in a working memory (or context) module in the prefrontal cortex. The models differ, however, concerning the scope of the attentional modulation, and as such complement each other. While in the model by Cohen et al. the attentional modulation selects a task (naming colors or naming words), in our case it helps selecting a target among distractors in a given modality. This subtle difference leads to an essential difference in the architecture. A task specification is diffuse in the sense that it acts on all objects related to that task (e.g., all words) and as such it can be relatively strong (since it does not give preference to one word over another). In contrast, feedback signals selecting a specific object (e.g., the letter "A") need to be much more subtle, lest they will falsely induce a state in the system that corresponds to the presence of a certain stimulus (e.g., an "A") in the absence of such input. On the other hand, this input needs to be effective enough to select the "A"-target when presented in the display. To obey both of these constraints, we assume that the attentional input is much weaker than the sensory input (the opposite is the case in the model by Cohen et al.). It is then due to recurrence and the lateral competition between cell assemblies that the small modulation determines the selection process.⁸ We expect that the actual attentional modulation will involve a combination of such strong diffuse and more subtle specific inputs.

Our model is distinguished from the four mentioned models by several features. One is that it is formulated explicitly in terms of neuronal units and populations. None of the other models uses neurally implausible mechanisms (as it was the case with some older models, many of them being formulated in the language of digital computers). Nevertheless, it remains to be shown to what extent these models can effectively be based on elements with the behavior of standard neurons. Another distinction between our model and the others is that the formulation of our model in terms of differential equations for the dynamics of its elements allows us to describe these dynamics in detail. This made it possible to compare the results of the model with electrophysiological recordings (Figs. 2-6) and make predictions about the response of individual neurons in terms of their detailed temporal dynamics.

A prediction of our model is the time course of neuronal activity in negative trials (trials in which the matching stimulus is not presented), which was not reported in the Chelazzi et al. (1993a) study. According to our model, we expect that, since the attentional modulation

is weaker than the sensory input, the activity of neurons in the distractor assembly will not show the fast suppression observed when the target is present (Fig. 4b). Instead, when several different distractors (and no target) are shown, we expect that due to the competition, the activity of these neurons will stay at an intermediate level, lower than for target cells in positive trials, but higher than for distractor cells when a target is present. Our model also makes predictions in the context of target-distractor similarity, in particular, larger responses for cells that respond to distractors similar to the target, than for responses to dissimilar distractors (see also Figs. 4 and 5).

In relation to visual performance, our model shows that a plausible neural mechanism can detect a target independent of the number of distractors. The degree to which the distractors can be discarded in parallel depends, however, on the spatial extent of the nonlinear inhibitory mechanism assumed in the preprocessing; if the spatial range of this inhibition is unlimited, completely parallel performance can be achieved. In contrast, for a limited inhibition range, set-size effects (Sagi & Julesz, 1987; Palmer et al., 1993; Palmer, 1994) are expected for low density displays. These effects should disappear when the element density becomes higher than the inhibition range.

We should emphasize that the neuronal models we are using are still very simple. For instance, the implementation of the top-down interactions is at present limited to linear additive synapses. More elaborated nonlinear synaptic interaction, e.g., based on NMDA receptors, may also play an important role in the neural mechanism responsible for top-down attentional modulation, and should be the subject of further research. Furthermore, in our model (as well as in the four other mentioned models), attentional selection is obtained through a parallel process. Recently, other models exploring the neuronal implementation of spatially defined selective attention based on serial processing have been developed by Niebur, Koch, and Rosin (1993), Olshausen, Anderson, and Van Essen (1993), and Niebur and Koch (1994). It remains a challenge for further experimental work to determine which of the various proposed processing modes of selective attention are realized, and in which situations. In particular, we expect that multi-electrode recordings in the awake, behaving monkey will shed light on the serial or parallel mechanisms underlying selective attention, and their dependence on the characteristics of the task.

APPENDIX: MODEL DYNAMICS

A The Integrate-and-Fire Scheme

Excitatory cells are divided in cell assemblies (or populations) that are either distinct (five populations of 200 cells each) or with small overlaps (of 5-10 cells). In the

latter case, the populations are arranged in a ring structure in which only nearest-neighbor populations share neurons, i.e., population “2” shares neurons with populations “1” and “3”, etc. Inhibition is modeled as a response from a common pool of neurons.

Each excitatory cell is simulated as a continuous membrane potential variable with the dynamics of an RC circuit (time constant $\tau = 20$ msec) with the addition of a reset mechanism. Once a cell reaches the threshold voltage, it emits a pulse that is transmitted in one time step (1 msec) to the other cells in the same assembly, and the potential is reset by subtracting the threshold voltage, V_{th} . The equation for the potential V_i of neuron i is as follows:

$$\tau \frac{dV_i}{dt} = -V_i + g_{exc}^i(t)[E_{exc} - V_i(t)] + I_i^{ext} - \beta I_{inh}(t) \quad (1)$$

for $V_i(t) < V_{th}$ and $V_i(t)$ is set to $V_i(t) - V_{th}$ if $V_i(t) > V_{th}$. The terms on the right-hand side of Eq. (1) are

- $g_{exc}^i(t)[E_{exc} - V_i(t)]$, the voltage change due to active cells in the same assembly, determined by the reversal potential of excitatory synapses E_{exc} , and by the synaptic conductance g_{exc}^i (in units of the leakage conductance of the cell), obtained from

$$g_{exc}^i(t) = \int_{-\infty}^t dt' e^{-t'/\tau_{exc}} \sum_j W_{ij}^E \theta(V_j - V_{th}) \quad (2)$$

where θ is the Heaviside function, $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ otherwise. The sum is over all excitatory cells, and $W_{ij}^E = W^E/N_{syn} > 0$ if there is a synapse from cell j to cell i and zero otherwise (Willshaw rule; see Willshaw, Buneman, & Longuet-Higgins, 1969). Connections are made by choosing N_{syn} postsynaptic cells randomly from all the members of the same assembly. No connections are made to cells in other assemblies, except for cells in the overlapping sets.

- I_i^{ext} , voltage changes due to external currents from three sources: a diffuse spontaneous background input originating in other brain areas, a sensory input, and a specific attentional input (only to the attended target assembly) from the working memory module. The three input sources generate Poisson processes of rates, λ_0 , λ_s , and λ_a , respectively. Each time any of these stochastic processes generates an event, a contribution $1/N_{syn}$ of the threshold is added to the membrane potential.
- βI_{inh} , where β has the units of voltage and is proportional to the inhibitory synaptic weights, and I_{inh} is the activity of the inhibitory pool obtained from the following differential equation, in which τ_i is a time constant and F a sigmoidal (logistic) function, $F(x) = \{1 + \exp[-(x - x_0)/\Delta]\}^{-1}$.

$$\tau_i \frac{dI_{inh}}{dt} = -I_{inh} + F\left[\sum_j \frac{1}{N} \theta(V_j - V_{th})\right] \quad (3)$$

The sum on the right-hand side of this equation runs again over all cells, representing thus the fraction of active neurons out of the total number of excitatory cells $N = 2000$.

The time constants used in the simulations are $\tau_{exc} = 2.8$ and $\tau_i = 1.4$, and the rates are $\lambda_0 = 2200$ Hz, $\lambda_s = 4200$ Hz, and $\lambda_a = 50$ Hz, respectively. The excitatory weights are $W^E = 0.35$ and the number of synapses per cell is $N_{syn} = 50$. The excitatory driving potential was $E_{exc} = 5$ in units of the threshold voltage (which was arbitrarily chosen as $V_{th} = 1$), and the inhibitory weight $\beta = 0.12$. The gain and threshold of the logistic function F are $x_0 = 0.03$ and $\Delta = 0.01$, respectively.

B The Mean-Field Population Dynamics Approach

In accordance with the Hebbian principle, stating that cognitive events are encoded by dedicated cell assemblies, it is compelling to develop a dynamic theory whose dependent variables are the activation levels of the relevant cell populations. While this is problematic in the general case (since a macroscopic description in terms of population activations contains less information than the underlying single cell neural description), such a reduction is possible under certain simplifying assumption, such as ergodicity and asynchronous activity. Under these assumptions, the discharge rate of a cell (averaged over some time window) is equivalent to the momentary activity of a neural population (replacing thus temporal with ensemble averaging).⁹

Dynamic equations for the activities of populations can be obtained from a mean-field approximation (Wilson & Cowan, 1972; Amit & Tsodyks, 1991; Abbott, 1992). Accordingly, each population i is characterized by two related variables: its activation x_i (equivalent to the average discharge rate), and an input current that is characteristic for all cells in the population, I_i , which satisfies the sigmoidal relation $x = F(I)$. Following Amit and Tsodyks (1991), we used

$$F(I) = \frac{1}{T_r - \tau \log[1 - 1/(\tau I)]} \quad (4)$$

which is the response function (transforming current into discharge rates) for an integrate-and-fire cell with deterministic input (where T_r is the cell's absolute refractory period, taken here as 1 msec, and τ is the membrane time constant). As opposed to the more familiar logistic sigmoids, frequently used in the connectionist literature, which are symmetric around the central value of the sigmoid, the integrate-and-fire response function

has a logarithmic singularity that reflects the threshold of the cell [given by $1/\tau$ in Eq. (4)]. The initial slope of the function is thus very steep and it gradually decreases to zero for large input currents. It was shown by Amit and Tsodyks (1991) that networks of neurons with this activation function (but not with logistic sigmoids) are able to sustain attractor states with low activity (less than 60 Hz), as they are observed in cortical areas related to working and sensory memory.

Taking into account the effect of noise (Gaussian; of width σ) in the input leads to a correction for small currents that smoothes the threshold singularity, since a signal of mean input smaller than the threshold will also evoke some response. The noise-corrected response function is given by $F(I, \sigma) = 1/[T_r + T_{sp}(I, \sigma)]$. The mean interspike interval T_{sp} in this equation is given by Ricciardi (1977)

$$T_{sp}(I, \sigma) = \tau \sqrt{\pi} \int_{x_1(I, \sigma)}^{x_2(I, \sigma)} \exp(z^2) [1 + \text{erf}(z)] dz \quad (5)$$

where erf is the error function and the limits of integration are

$$x_1(I, \sigma) = -\frac{I\tau}{\sigma\sqrt{\tau}}$$

$$x_2(I, \sigma) = \frac{1 - I\tau}{\sigma\sqrt{\tau}}$$

The response function in the presence and absence of noise is displayed in Figure 7. As shown in Amit and Tsodyks (1991), the noise correction preserves the ability to sustain low activity attractor states.

The current equations for population i are then

$$-\tau_s \frac{dI_i}{dt} = -I_i + AF(I_i) + A_2 [F(I_{i-1}) + F(I_{i+1})] - BF(I^b) + I_i^{\text{ext}} \quad (6)$$

$$-\tau_s \frac{dI_b}{dt} = -I^b + C \sum_i F(I_i) - DF(I^b) \quad (7)$$

where I_i is the input current for population i and I^b is the current in the inhibitory population. The parameters A, B, C, D represent the synaptic weights ($A = W^E \tau_{\text{exc}} = 0.95, B = 0.8 = \beta, C = 1.0, D = 0.1$). $A_2 = 0.15$ is the lateral excitatory weight, between similar (neighboring) assemblies in the ring structure. The synaptic time constants were $\tau_s = 5$ msec, and the membrane time constant $\tau = 20$ msec (this determines the threshold current as $1/\tau = 0.05$). The external current $I_i^{\text{ext}} = I_i^0 + I_i^s + I_i^q + R_i$ is composed of four components:

- I_i^0 —a diffuse spontaneous background input (same for all neurons in the network and constant during the simulation, $I_i^0 = 0.025$).
- I_i^s —a sensory input to the cells belonging to the target

assembly for $t < 300$ msec and $700 < t < 1000$ msec and on the distractor assembly for $700 < t < 1000$ msec ($I_i^s = 0.05$).

- I_i^q —an attentional input to the target assembly, during the whole simulation ($I_i^q = 0.005$).
- R_i —a Gaussian noise term of zero mean and $SD = \sigma$, which reflects input fluctuations that are not averaged out in the population averaging ($\sigma = 0.03$ current units per msec).

Equations (6)–(7) were solved numerically, and the activity variables $x_i(t) = F[I_i(t)]$, are displayed in Figures 4–6.

An understanding of the behavior of the system can be obtained by examining its fixed points. Although the fixed point analysis for the general system in Eqs. (6)–(7) is rather complex, a qualitative understanding can be obtained by neglecting the inhibition and the lateral coupling A_2 . Under such simplifications (which are justified during the delay interval), the fixed points of Eqs. (6)–(7) are characterized by only one nonlinear equation determined by the response function. In Figure 8 we display the corresponding fixed points obtained with the deterministic response function, for three values of input, corresponding to

- $I_i^{\text{ext}} = 0$ (the rightmost curve). The only fixed point is at zero activation.
- $I_i^{\text{ext}} = 0.6/\tau = 0.03$ (the central curve). This corresponds to the input during the delay period in the attended assembly. The almost tangent relationship leads to the very slow decay of activity during the delay period, in the attended assembly.
- $I_i^{\text{ext}} = 1.25/\tau = 0.075$ corresponding to the input during the sensory stimulation (leftmost curve). The resulting fixed point has a much larger activation. Note, however, that while the simplification of vanishing inhibition was justified in the previous cases, this is not true anymore during the period of sensory input. In this condition, the actual fixed point is lower due to the competitive inhibition.

C Preprocessing—Nonlinear Inhibition

The preprocessing stage implements a nonmonotonic response function that accounts for the fact that two identical shapes in the display inhibit each other, providing a weaker input to the corresponding assembly in inferotemporal cortex as observed in Miller et al. (1993b). We assume that in this preprocessing stage, the activity of each cell assembly is regulated by shunting inhibition. As shown by Abbott (1991), shunting inhibition near the cell soma (where most of the inhibitory synapses are observed) leads to a nonlinear (exponential) damping of the cell's current, due to an increase in the membrane conductance and shortening of the cell's electrotonic length. Assuming that local excitatory and inhibitory cells receive an input proportional to the

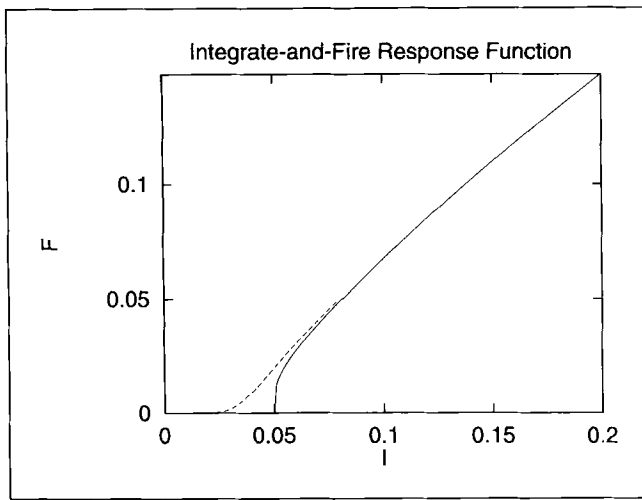


Figure 7. Response function of integrate-and-fire neurons. The solid curve corresponds to Eq. (4) with parameters $T_r = 1$ msec, $\tau = 20$ msec, and the dashed curve corresponds to Eq. (5) with the same T_r and τ , and $\sigma = 0.05$. The influence of noise is significant only for small input currents where it smooths the logarithmic singularity.

number n of identical shapes, the output O of the excitatory preprocessing unit (which provides input to the assemblies in the sensory memory module) decreases in our model for increasing value of $n > 1$.

$$O = An \exp(-B\sqrt{n}) \quad (8)$$

The constants A and B are chosen in our simulation as $A = 0.41$ and $B = 2.2$.

For instance, the input to the sensory memory was $O(n = 1) = 0.045$ when one shape of a given type was presented, and $O(n = 2) = 0.037$ when two identical shapes were presented.

The local inhibitory mechanism is expected to work not only on perfectly identical targets, but also on similar targets (in proportion to their degree of similarity). We implemented the inhibition of "similar" shapes by assuming a fan-out from the input layer to the preprocessing layer. Each input unit projects to the corresponding unit in the preprocessing layer above it with a weight of 1 and to the neighboring units with weights $\lambda = 0.25$.

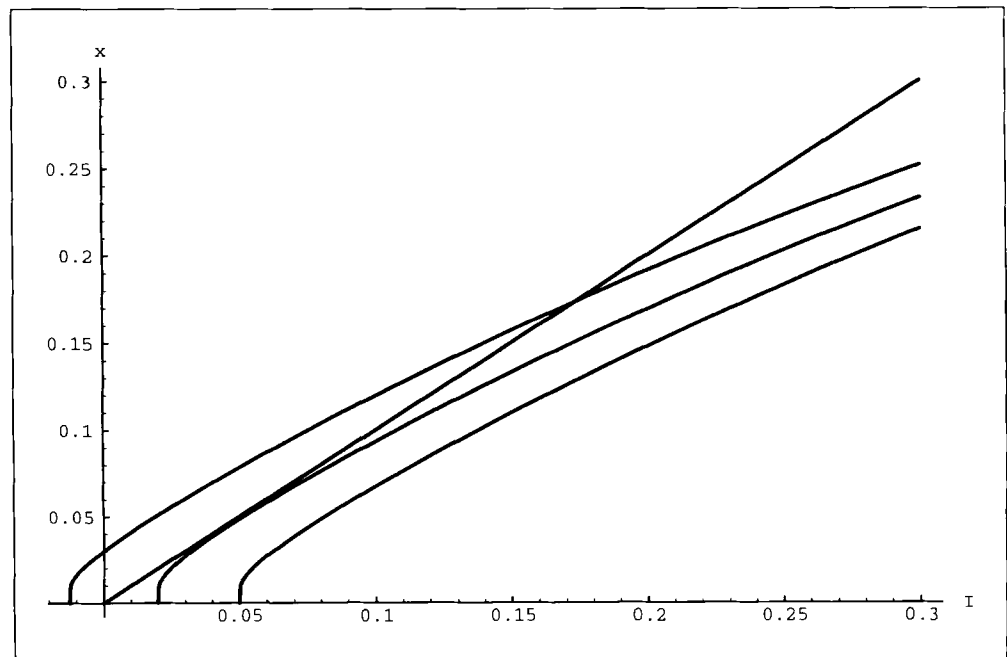
Consequently, two similar stimuli in the input layer (i.e., neighboring units on the ring) will send an input corresponding to $n = 1.25$ in Eq. (8) to the corresponding excitatory and inhibitory preprocessing units. In this case, the corresponding input to the sensory layer is $O(1.25) = 0.043$. The small difference between this input and the one obtained by the sensory memory module for nonsimilar shapes (0.045) leads to the distractor similarity effect shown in Figure 6.

Acknowledgments

M.U. was supported by the National Institute of Mental Health (Grants MH47566 and MH45156) and E.N. by grants from the NSF, AFOSR, and ONR to Christof Koch. We wish to thank Jonathan Cohen, Christof Koch, and James McClelland for insightful discussions and support, Jochen Braun for a critical reading of the manuscript, and Bob Desimone and Leonardo Chelazzi for stimulating comments.

Reprint requests should be sent to Ernst Niebur, Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA.

Figure 8. Fixed point solution of Eqs. (6)–(7) in the absence of inhibition and lateral coupling for three different values of external current I_i^{ext} , corresponding to sensory input (leftmost curve), top-down stimulation during the delay period for the attended stimulus (central curve), and no input (rightmost curve), respectively. Fixed points are the intersections of the identity function (straight line) with the response curves.



Notes

1. This is a simplification. Wolfe (1992) showed that although "effortless" texture segregation in figure-ground segregation tasks is in general correlated with parallel visual search, this is not always the case. Certain stimuli that lead to effortless texture segregation do not pop out in visual search tasks, and vice versa.
2. It has also been reported recently that originally serial tasks can become parallel with training for some observers (Sireteanu & Rettenbach, 1994).
3. In the connectionist literature the ART model (Carpenter & Grossberg, 1987) uses a similar principle, based on competing nodes for categorization.
4. This condition can be obtained if the self-excitation coefficient, A (see Appendix B), which regulates the maximal slope of the response function, is smaller than, but close to, unity. This coefficient was chosen as $A = 0.95$ in Appendix B and corresponds to $W^E = A/\tau_{exc}$ in Appendix A. For this choice, no fixed-point solution corresponding to a sustained activity state is possible in the absence of sensory input (see Fig. 8).
5. This is easily achieved by choosing a self-excitation coefficient A greater than unity.
6. Shifting the response curve close to a tangency relationship of the identity (see Appendix B).
7. Some psychophysics experiments indicate that the inhibitory interaction among like stimuli is spatially limited (Sagi & Julesz, 1987); thus target selection is easier only if the distractors are in spatial proximity. In this work, for the sake of simplicity, we do not address interactions based on such spatial relations.
8. An alternative could be to model attentional top-down influence as a multiplicative rather than an additive modulation.
9. In the presence of synchronized activity, single cell discharge rates are not equivalent with the population activity. The reason is that while the first may be constant the second may oscillate. In our model the population operates in an asynchronous mode, allowing, therefore, the assumption of ergodicity to be made.

REFERENCES

- Abbott, L. (1991). Realistic synaptic inputs for model neural networks. *Network*, 2, 245-258.
- Abbott, L. (1992). Firing rate models for neural populations. In O. Benhar, C. Bosio, P. Giudice, & E. Tabet (Eds.), *Neural networks: From biology to high energy physics*. Pisa: ETS Editrice.
- Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14, 105-126.
- Amit, D., & Tsodyks, M. (1991). Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate-spikes, rates and neuronal gain. *Network*, 2, 259-273.
- Bergen, J., & Julesz, B. (1983). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13, 857-863.
- Bernander, O., Koch, C., & Usher, M. (1994). The effect of synchronized inputs at the single neuron level. *Neural Computation*, 6(4), 622-641.
- Blakemore, C., & Tobin, E. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, 15, 439-446.
- Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception and Psychophysics*, 48, 45-58.
- Braun, J., & Sagi, D. (1991). Texture-based tasks are little affected by second tasks requiring peripheral or central attentive fixation. *Perception*, 20, 483-500.
- Carpenter, G., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54-115.
- Cave, K., & Wolfe, J. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.
- Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (1993a). A neural basis for visual search in inferior temporal cortex. *Nature (London)*, 363, 345-347.
- Chelazzi, L., Miller, E., Lueschow, A., & Desimone, R. (1993b). Dual mechanisms of short-term memory: ventral prefrontal cortex. *Society for Neuroscience Abstracts*, 19, 975.
- Cohen, J., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes. *Psychological Review*, 97, 332-361.
- Cohen, J., Servan-Schreiber, D., & McClelland, J. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology*, 105, 239-269.
- Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Science, USA*, 81, 4586-4590.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272-300.
- Duncan, J. (1989). Boundary conditions on parallel processing in human vision. *Perception*, 18, 457-469.
- Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458.
- Egeth, H., Virzi, R., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology*, 10(1), 32-39.
- Eriksen, C. W., & Hoffman, J. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception and Psychophysics*, 14, 155-160.
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40, 225-240.
- Eriksen, C., & Yeh, Y. (1985). Allocation of attention in the visual field. *Human Perception and Performance*, 11, 583-597.
- Fuster, J. M., & Jervey, J. (1980). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*, 212, 952-955.
- Georgopoulos, A., Kettner, R., & Schwartz, A. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *Journal of Neuroscience*, 8, 2928-2937.
- Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Krieger.
- Grossberg, S., Mingolla, E., & Ross, W. (1994). A neural theory of attentive visual search: interactions at boundary, surface, spatial and object recognition. *Psychological Review*, 101(3), 470-489.
- Heit, G., Smith, M., & Halgren, E. (1988). Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature (London)*, 333, 773-335.
- Helmholtz, H. v. (1867). *Handbuch der physiologischen Optik*. Leipzig: Voss.
- Hendin, O., Horn, D., & Usher, M. (1991). Chaotic behavior of a neural network with dynamical thresholds. *International Journal of Neural Systems*, 1(4), 327-335.
- Horn, D., & Usher, M. (1990). Excitatory-inhibitory networks with dynamical thresholds. *International Journal of Neural Systems*, 1(3), 249-257.
- Humphreys, G., & Müller, H. (1993). Search via recursive re-

- jection (SERR): A connectionist model of visual search. *Cognitive Psychology*, 25, 43–110.
- James, W. (1890). *The principles of psychology*. New York: Henry Holt.
- Jonides, J. (1980). Toward a model of the mind's eye. *Canadian Journal of Psychology*, 34, 103–112.
- Jonides, J. (1983). Further toward a model of the mind's eye movement. *Bulletin of the Psychonomic Society*, 21, 247–250.
- Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, 55, 289–298.
- Krose, B., & Julesz, B. (1989). The control and speed of shifts of attention. *Vision Research*, 29, 1607–1619.
- McClelland, J., Rumelhart, D., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.
- Miller, E., Gochin, P., & Gross, C. (1993a). Suppression of visual responses of neurons in inferior temporal cortex of the awake monkey by addition of a second stimulus. *Brain Research*, 616, 25–29.
- Miller, E., Li, L., & Desimone, R. (1993b). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience*, 13, 1460–1478.
- Miyashita, Y., & Chang, H. (1988). Neural correlate of pictorial short term memory in primate temporal cortex. *Nature (London)*, 331, 68–70.
- Motter, B. (1994a). Neural correlates of attentive selection for colors or luminance in extrastriate area V4. *Journal of Neuroscience*, 14, 2178–2189.
- Motter, B. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience*, 14, 2190–2199.
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, 29, 1631–1647.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Niebur, E., & Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, 1(1), 141–158.
- Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neural basis of attention. *Vision Research*, 33, 2789–2802.
- Olshausen, B., Andersen, C., & Van Essen, D. (1993). A neural model of visual attention and invariant pattern recognition. *Journal of Neuroscience*, 13(11), 4700–4719.
- Palmer, J. (1994). Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vision Research*, 34, 1703–1721.
- Palmer, J., Ames, C., & Lindsey, D. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 108–130.
- Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- Ricciardi, L. (1977). *Diffusion processes and related problems*. Berlin: Springer-Verlag.
- Rumelhart, D., McClelland, J., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Sagi, D., & Julesz, B. (1985). Where and what in vision. *Science*, 228, 1217–1219.
- Sagi, D., & Julesz, B. (1987). Short range limitations on detection of features differences. *Spatial Vision*, 2, 39–49.
- Shaw, M. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 586–598.
- Shaw, M., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 201–211.
- Sireteanu, R., & Rettenbach, R. (1994). Serial visual search can become parallel with practice. *Investigative Ophthalmology and Visual Science*, 35(4), 2080.
- Softky, W., & Koch, C. (1993). The highly irregular firing of cortical-cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1), 334–350.
- Tanaka, K., & Saito, Y. (1991). Coding visual images of objects in the inferotemporal cortex of macaque monkey. *Journal of Neurophysiology*, 66, 170–189.
- Townsend, J. (1972). Some results concerning the identifiability of parallel and serial search. *British Journal of Mathematical and Statistical Psychology*, 25, 168–199.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194–214.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Human Perception and Performance*, 16, 459–478.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141.
- Tsal, Y., & Lavie, N. (1988). Attending to color and shape: The special role of location in selective processing. *Perception and Psychophysics*, 44(1), 15–21.
- Tsal, Y., & Lavie, N. (1993). Location dominance in attending to color and shape. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 131–139.
- Vecera, P., & Farah, J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, 123, 146–160.
- Willshaw, D., Buneman, O., & Longuet-Higgins, H. (1969). Non-holographic associative memory. *Nature (London)*, 222.
- Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biological Cybernetics*, 12, 1–24.
- Wolfe, J. (1992). Effortless texture segmentation and parallel visual search are not the same thing. *Vision Research*, 32(4), 757–763.
- Wolfe, J., Cave, K., & Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology*, 15, 419–433.